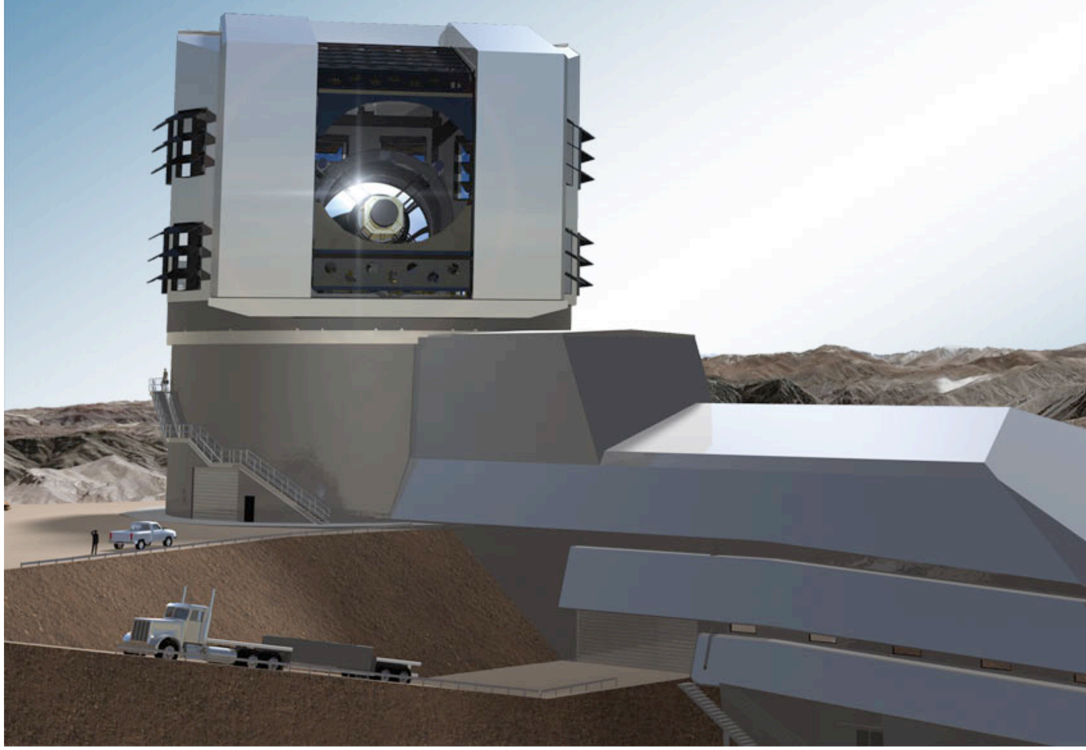
Photometric Redshifts: Potential Systematics and Solutions

Jeffrey Newman, U. Pittsburgh / PITT-PACC

Analysis Coordinator and acting Photo-z Working Group convener, LSST Dark Energy Science Collaboration

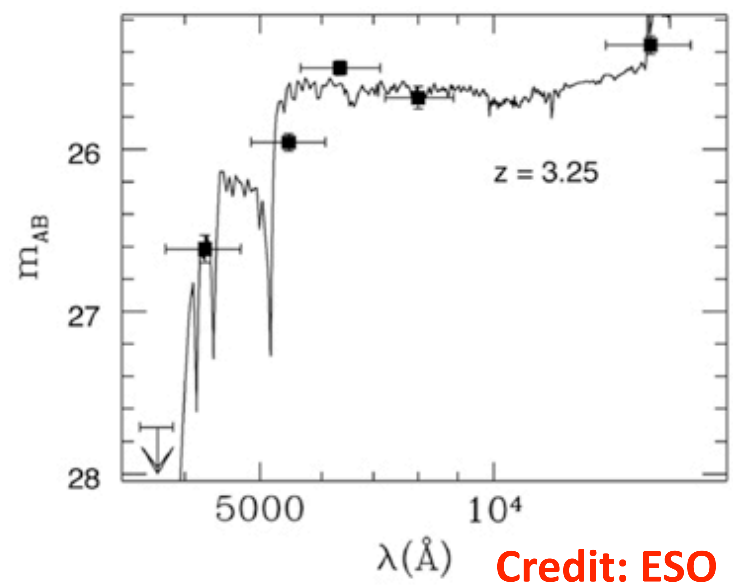


- **See Snowmass white papers on *Cross-Correlations* and *Spectroscopic Needs for Imaging Dark Energy Experiments* for all details!**

Spectroscopy provides ideal redshift measurements – but is infeasible for large samples

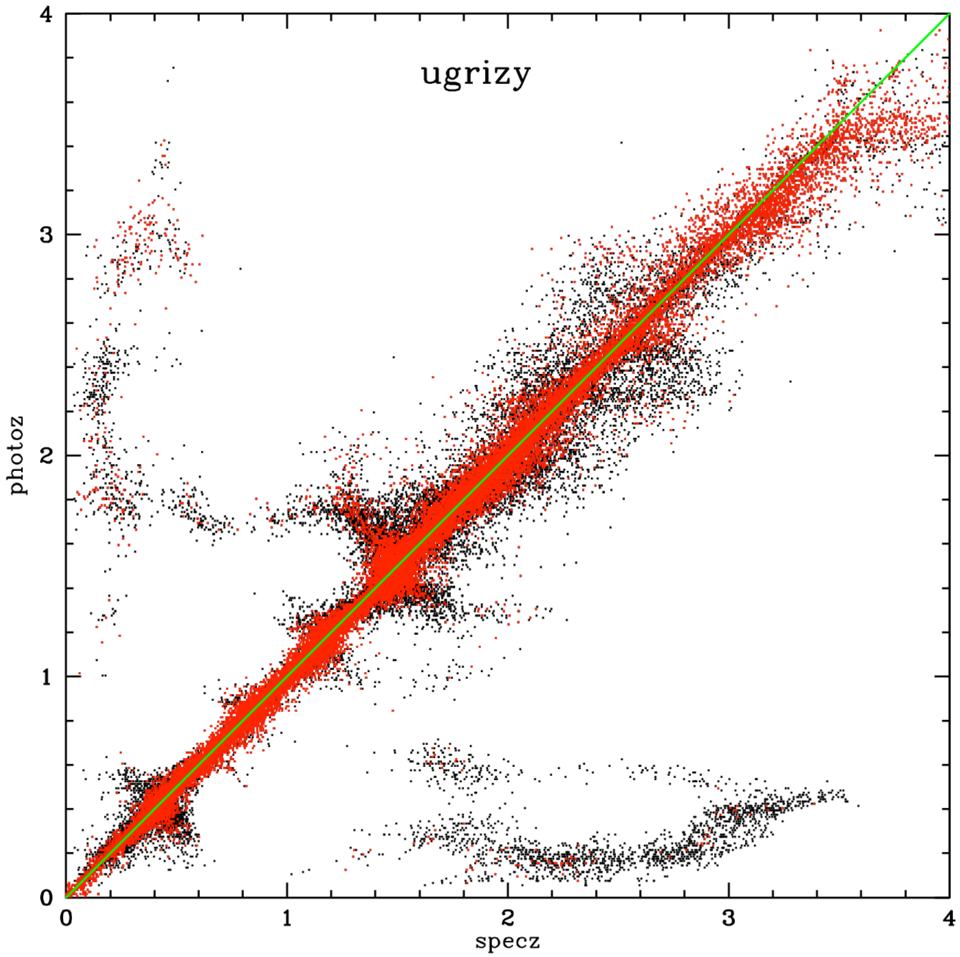
- For most dark energy probes, we wish to determine the dependence of some observable on redshift, *z*
- At LSST depths (i<25.3), ~190 hours on a 10m telescope to determine a redshift (~75% of time) spectroscopically
- With a next-generation, 5000-fiber spectrograph would take >50,000 10m telescope-years to measure redshifts for LSST "gold" weak lensing sample (4 billion galaxies)!

- Alternative: use broad spectral features to determine *z* : a photometric redshift
- Advantage: high multiplexing
- Disadvantages: lower precision, calibration uncertainties

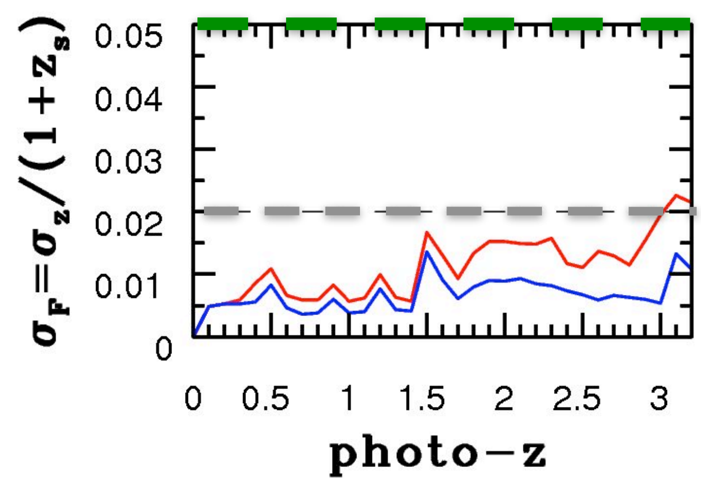


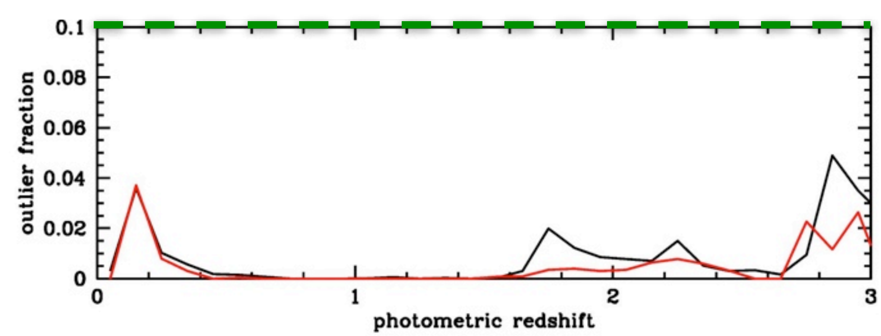
Credit: ESO

Example: expected photo-z performance for LSST *ugrizy*



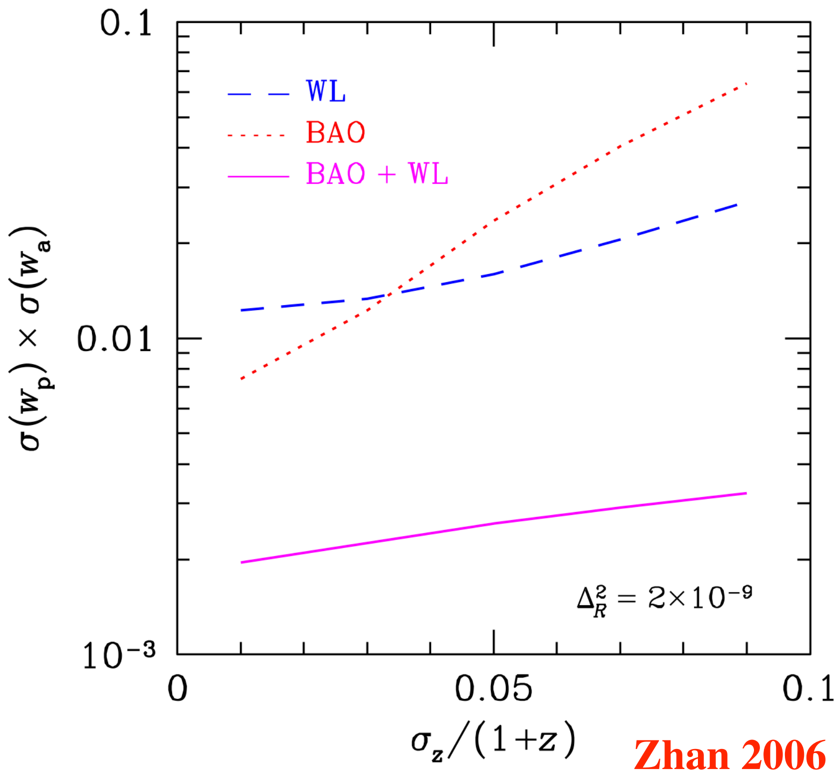
Green: Requirements on actual performance; grey: requirements on performance <u>with perfect template knowledge</u> (as in these sims)





Two spectroscopic needs for photo-z work: training and calibration

- **Better training of algorithms using objects with spectroscopic redshift measurements shrinks photo-z errors and improves DE constraints, esp. for BAO and clusters**
  - **Training datasets will contribute to calibration of photo-z's. ~Perfect training sets can solve calibration needs.**

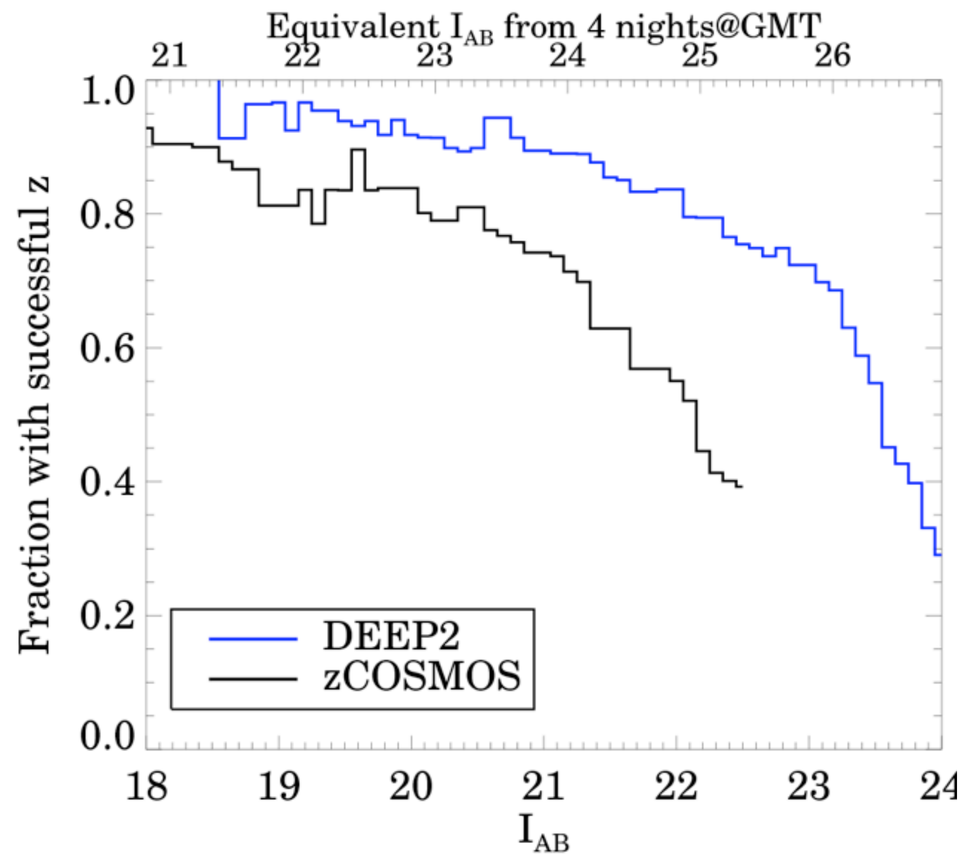


Zhan 2006

- **For weak lensing and supernovae, individual-object photo-z's do not need high precision, but the calibration must be accurate - bias and errors need to be extremely well-understood**
  - ***uncertainty in* bias, $\sigma(\delta_z) = \sigma(<z_p - z_s>)$, and in scatter, $\sigma(\sigma_z) = \sigma(\mathrm{RMS}(z_p - z_s))$, must both be <~0.002(1+*z*) for Stage IV**

Biggest concern: incompleteness in training sets

- In current deep redshift surveys (to i~22.5/R~24), 25-60% of targets fail to yield secure (>95% confidence) redshifts
- Redshift success rate depends on galaxy properties - losses are systematic, not random
- Estimated need 99-99.9% completeness to prevent systematic errors in calibration from missed populations

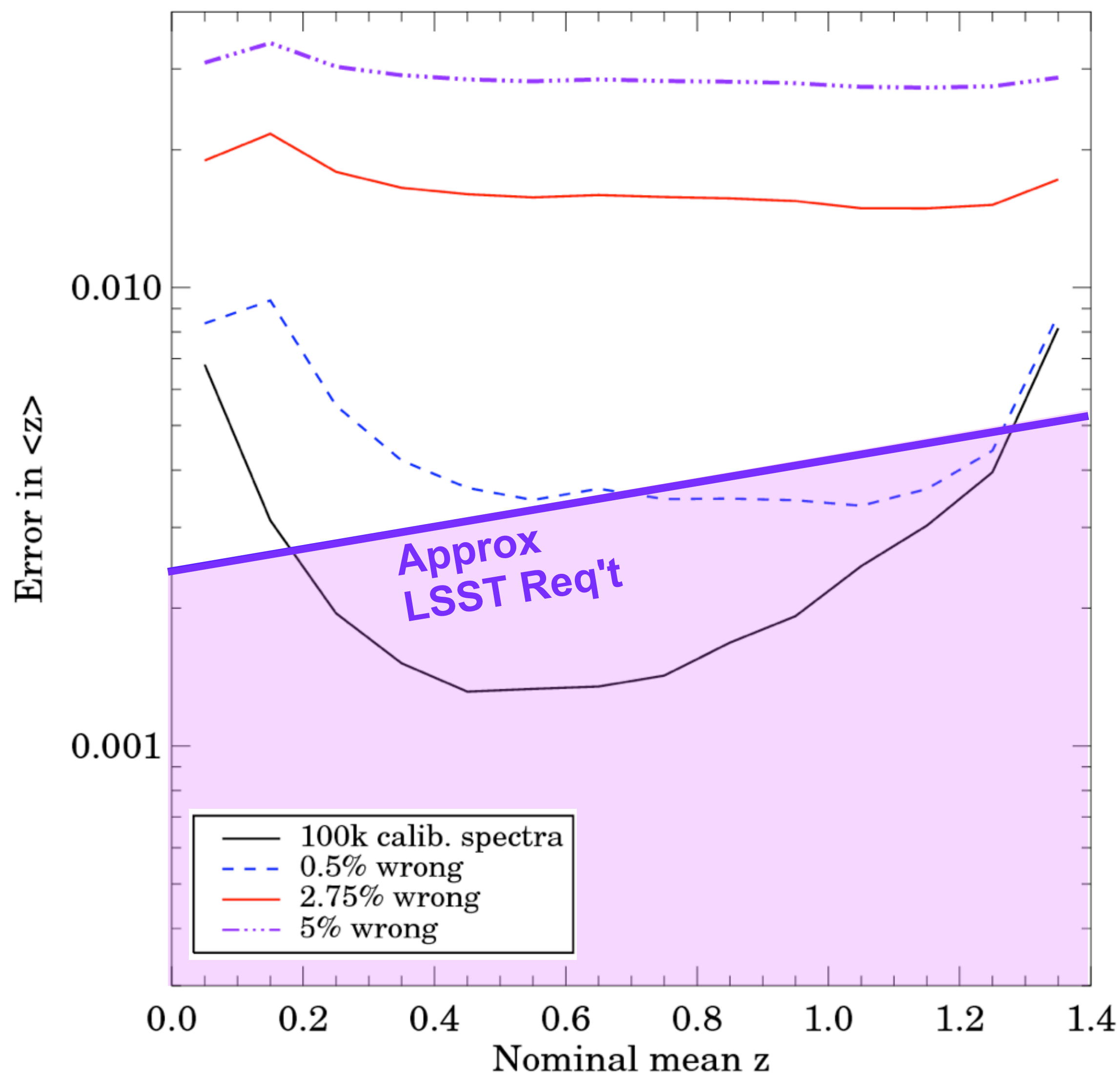


Data from DEEP2 (Newman et al. 2013) and zCOSMOS (Lilly et al. 2009)

Note: even for 100% complete samples, current false-z rates would compromise calibration accuracy

- Only the highest-confidence redshifts should be useful for precision calibration: lowers spectroscopic completeness further when restrict to only the best

Based on simulated redshift distributions for ANNz-defined DES bins in mock catalog from Huan Lin, UCL & U Chicago, provided by Jim Annis



What qualities do we desire in training spectroscopy?

- Sensitive spectroscopy of ~30,000 faint objects (to *i*=25.3 for LSST)
  - Needs a combination of large aperture and long exposure times
- High multiplexing
  - Required to get large numbers of spectra
- Coverage of full ground-based spectral window
  - Ideally, from below 4000 Å to ~1.5μm
- Significant resolution ($R=\lambda/\Delta\lambda>\sim4000$) at red end
  - Allows secure redshifts from [OII] 3727 Å line at $z>1$
- Field diameters > ~20 arcmin
  - Need to span several correlation lengths for accurate clustering
- Many fields, >~15
  - To mitigate sample/cosmic variance)

Estimated time requirements for training sets

- DES / 75% complete:
  ➡ 0.05 - 0.45 years (c. 2018), 0.02+ years (c. 2022+)
- DES / 90% complete:
  ➡ 0.34 - 1.6 years (c. 2018), 0.13 years (c. 2022+)
- LSST / 75% complete:
  ➡ 1.1 - 5.1 years (c. 2018), 0.42+ years (c. 2022+)
- LSST / 90% complete:
  ➡ 6.9 - 32 years (c. 2018), 2.6+ years (c. 2022+)

Depending on telescope/spectrograph properties, time required is determined by # of fields (15), # of spectra observable simultaneously (if multiplexing is low), or telescope field of view (if <<20' diameter). See Tables 2-1 & 2-2 of white paper.

3 Ways to address spectroscopic incompleteness – all may be feasible

I. Throw out objects lacking secure photo-*z* calibration
   - ID regions in e.g. *ugrizy* space where redshift failures occurred
   - Eliminating a fraction of sample has modest effect on FoM

     - Not yet known if sufficiently clean regions exist

II. Incorporate additional information
   - Longer exposure/wider wavelength range spectroscopy (JWST, etc.) for objects that fail to give redshifts in first try

     - Not yet known if will yield sufficient completeness

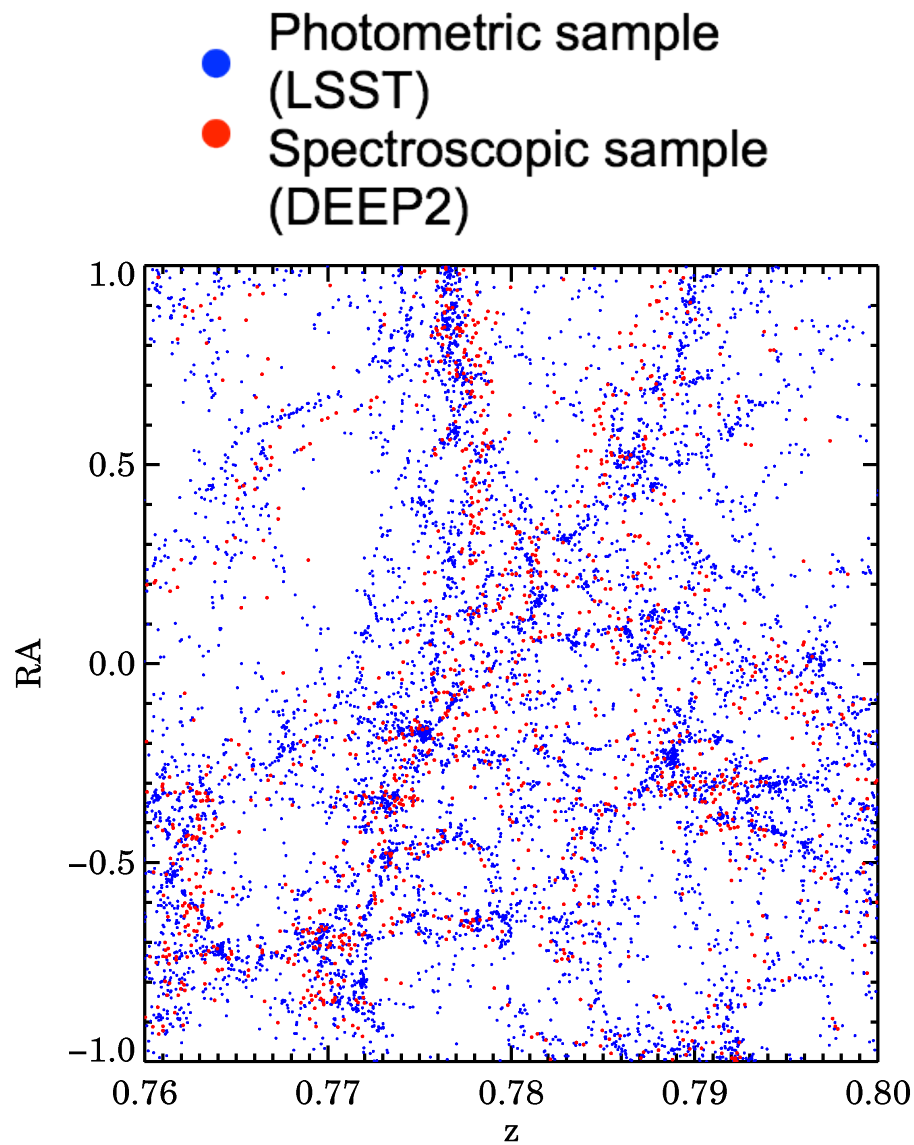
   - Develop comprehensive model of galaxy spectral evolution constrained by redshifts obtained

     - A major research program, not there now

III. Cross-correlation techniques

Cross-correlation methods for photo-z calibration

- **Galaxies of all types cluster together: trace same dark matter distribution**
- **Galaxies at significantly different redshifts do not cluster together**
- **Using observed clustering of objects in one sample vs. another, can determine the fraction of objects in overlapping redshift range**
- **Do this as a function of spectroscopic *z* to recover *p(z)***

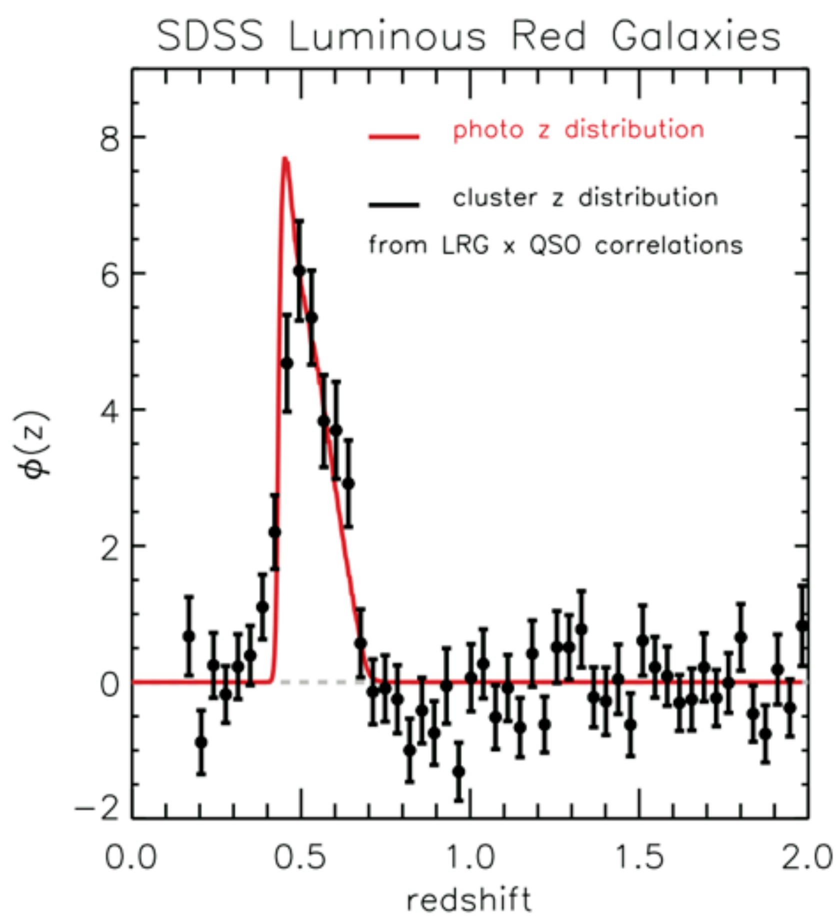


Higher-resolution information can be obtained by cross-correlating with spectroscopic samples

- Key advantage: spectroscopic sample can be systematically incomplete and include only bright galaxies!
- See: Newman 2008, Matthews & Newman 2010, 2011

Red: Photo-z distribution for LRGs in SDSS

Black: Cross-correlation reconstruction using only SDSS QSOs (rare at low z!)



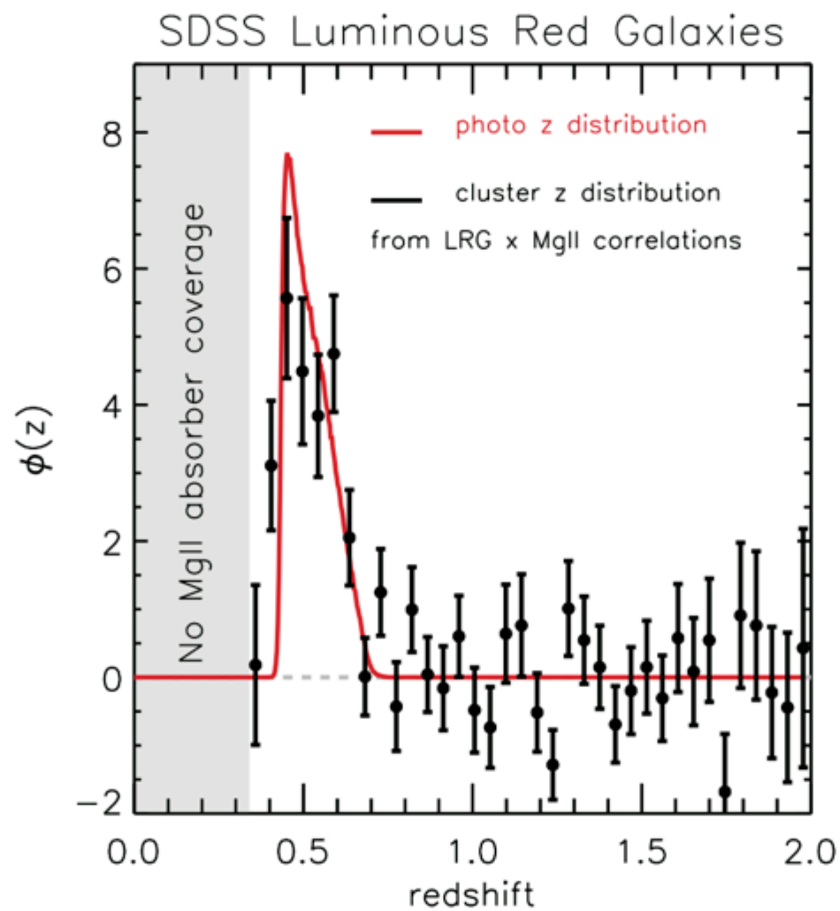
Menard et al. 2013

Higher-resolution information can be obtained by cross-correlating with spectroscopic samples

- **Key advantage: spectroscopic sample can be systematically incomplete and include only bright galaxies!**
- **See: Newman 2008, Matthews & Newman 2010, 2011**

Red: Photo-z distribution for LRGs in SDSS

Black: Cross-correlation reconstruction using only SDSS Mg II absorbers (even rarer!)



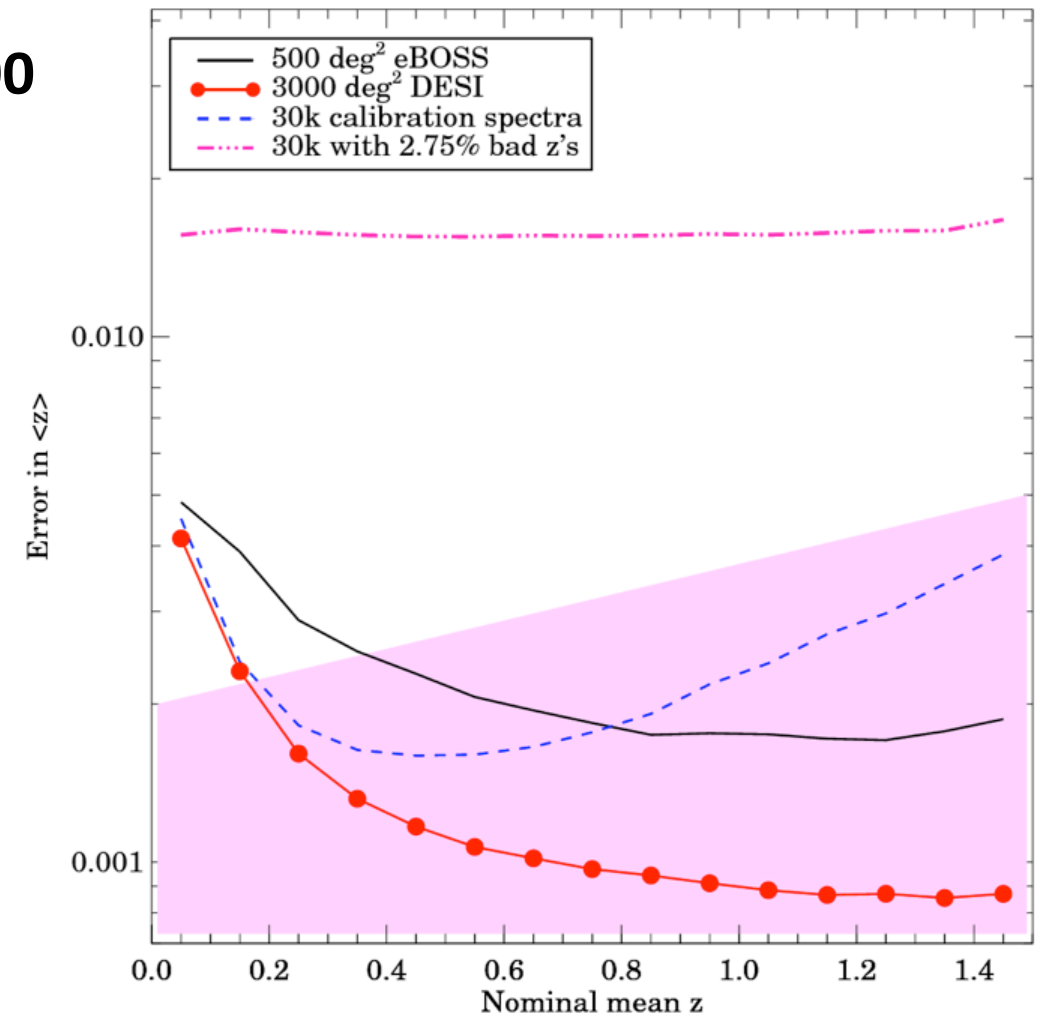
Menard et al. 2013

Spectroscopic requirements for cross-correlation methods

- Want >100k objects over >100 sq. degrees, spanning redshift range of photometric sample

- >500 square degrees of overlap with DESI-like survey sufficient for cross-correlation calibrations to Stage IV requirements

- Expected ~3000 deg^2 overlap is comparable to 100% complete sample of 100k spectra with no false z's!



Snowmass White Paper: Spectroscopic Needs for Imaging DE Experiments

Conclusions

- Photo-z's are critical for dark energy experiments
- Incompleteness or incorrect redshifts in spectroscopic samples will cause systematic errors in photo-z applications
- Cross-correlation methods can calibrate photometric redshifts even using incomplete samples of only bright galaxies & QSOs
- Minimum LSST photo-z training survey, ~75% complete:
  - 15 widely-separated pointings, ~30,000 spectra to $i = 25.3$, ~0.4 years on a 20-40m telescope (can do galaxy evolution science simultaneously)
- eBOSS and especially DESI are extremely useful for cross-correlation calibration

- See the Snowmass white papers on *Cross-Correlations* and *Spectroscopic Needs for Imaging Dark Energy Experiments* for much more!

Two basic Photo-z methods: Template fitting and training-based

- **Template fitting**: use galaxies with known *z* to determine set of underlying galaxy spectral energy distributions (SEDs) and relative photometric calibrations
  - Can then determine *p(z | ugrizy)*
  - For high accuracy, needs spectra of galaxies spanning full range of possible properties

- **Training-based:** Use set of galaxies with known redshift and well-understood sampling to determine relations between *z* and colors
- Training set MUST span full range of properties & z of galaxies
  - Pro: Takes advantage of progress in machine learning & stats
  - Con: Sensitive to systematic incompleteness in training sets - extrapolate poorly

Spectroscopic training set requirements

- Goal: make δ_z and $\sigma(\sigma_z)$ so small that systematics are subdominant
- Many estimates of training set requirements (Ma et al. 2006, Bernstein & Huterer 2009, Hearin et al. 2010, LSST Science Book, etc.)
- General consensus that roughly 20k-30k extremely faint galaxy spectra are required to characterize:
  - Typical z_{spec}-z_{phot} error distribution
  - Accurate catastrophic failure rates for all objects with $z_{phot} < 2.5$
  - Characterize all outlier islands in z_{spec}-z_{phot} plane via targeted campaign (core errors easier to determine)
- Those numbers of redshifts are achievable even at LSST depths, but. . .

Summary of potential instruments

Telescope / Instrument	Collecting Area (m^2)	Field area ($arcmin^2$)	Multiplex	Limiting factor
Keck / DEIMOS	76	54.25	150	Multiplexing
VLT / MOONS	58	500	500	Multiplexing
Subaru / PFS	53	4800	2400	# of fields
Mayall 4m / DESI	11.4	25500	5000	# of fields
WHT / WEAVE	13	11300	1000	Multiplexing
GMT/MANIFEST+GMACS	368	314	420-760	Multiplexing
TMT / WFOS	655	40	100	Multiplexing
E-ELT / OPTIMOS	978	39-46	160-240	Multiplexing

Table 2-1. *Characteristics of current and anticipated telescope/instrument combinations relevant for obtaining photometric redshift training samples. Assuming that we wish for a survey of ~15 fields of at least 0.09 deg^2 each yielding a total of at least 30,000 spectra, we also list what the limiting factor that will determine total observation time is for each combination: the multiplexing (number of spectra observed simultaneously); the total number of fields to be surveyed; or the field of view of the selected instrument. For GMT/MANIFEST+GMACS and VLT/OPTIMOS, a number of design decisions have not yet been finalized, so a range based on scenarios currently being considered is given.*

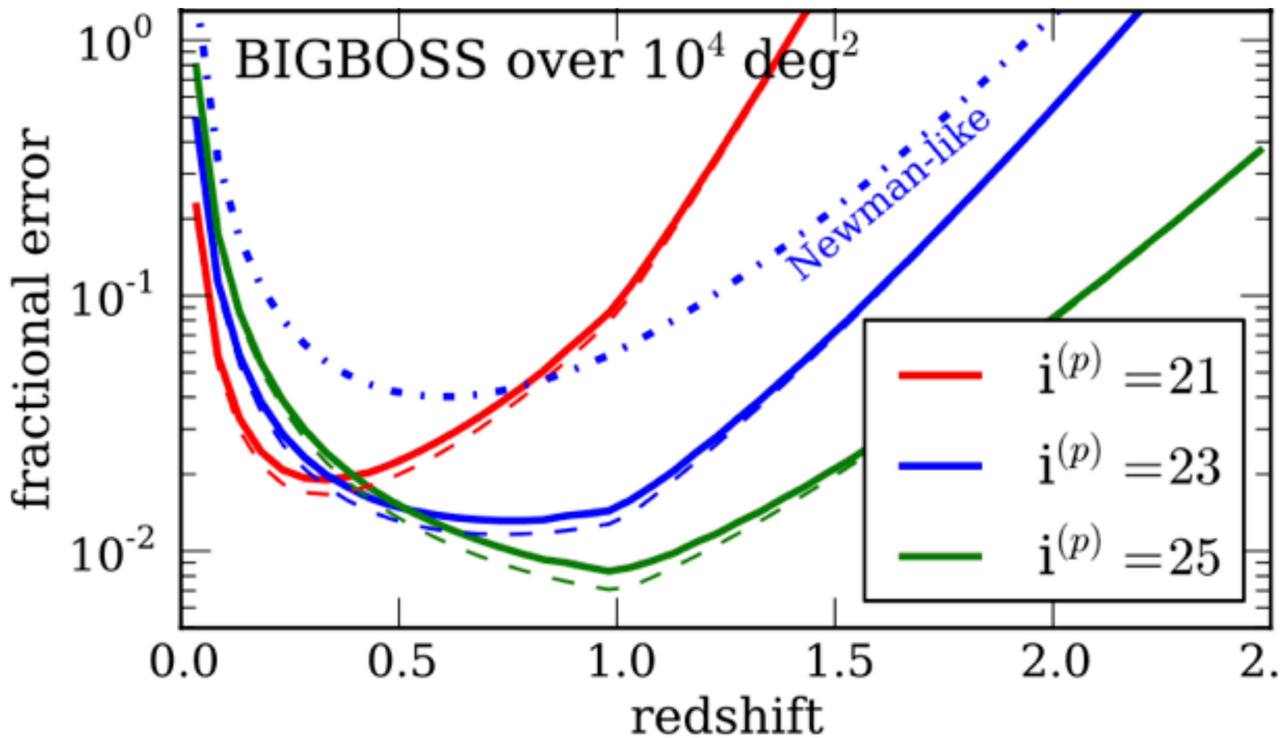
Time required for each instrument

Telescope / Instrument	Total time(y), DES / 75% complete	Total time(y), LSST / 75% complete	Total time(y), DES / 90% complete	Total time(y), LSST / 90% complete
Keck / DEIMOS	0.51	10.22	3.19	63.89
VLT / MOONS	0.20	4.00	1.25	25.03
Subaru / PFS	0.05	1.10	0.34	6.87
Mayall 4m / DESI	0.26	5.11	1.60	31.95
WHT / WEAVE	0.45	8.96	2.80	56.03
GMT/MANIFEST+GMACS	0.02 - 0.04	0.42 - 0.75	0.13 - 0.24	2.60 - 4.71
TMT / WFOS	0.09	1.78	0.56	11.12
E-ELT / OPTIMOS	0.02 - 0.04	0.50 - 0.74	0.16 – 0.23	3.10 - 4.65

Table 2-2. *Estimates of required total survey time for a variety of current and anticipated telescope/instrument combinations relevant for obtaining photometric redshift training samples. Calculations assume that we wish for a survey of ~15 fields of at least 0.09 deg^2 each, yielding a total of at least 30,000 spectra. Survey time depends on both the desired depth (i=23.7 for DES, i=25.3 for LSST) and completeness (75% and 90% are considered here). Exposure times are estimated by requiring equivalent signal-to-noise to 1-hour Keck/DEIMOS spectroscopy at i~22.5. GMT / MANIFEST + GMACS estimates assume that the full optical window may be covered simultaneously at sufficiently high spectral resolution; in some design scenarios currently being considered, that would not be the case, increasing required time accordingly.*

Previous cross-correlation forecasts are pessimistic!

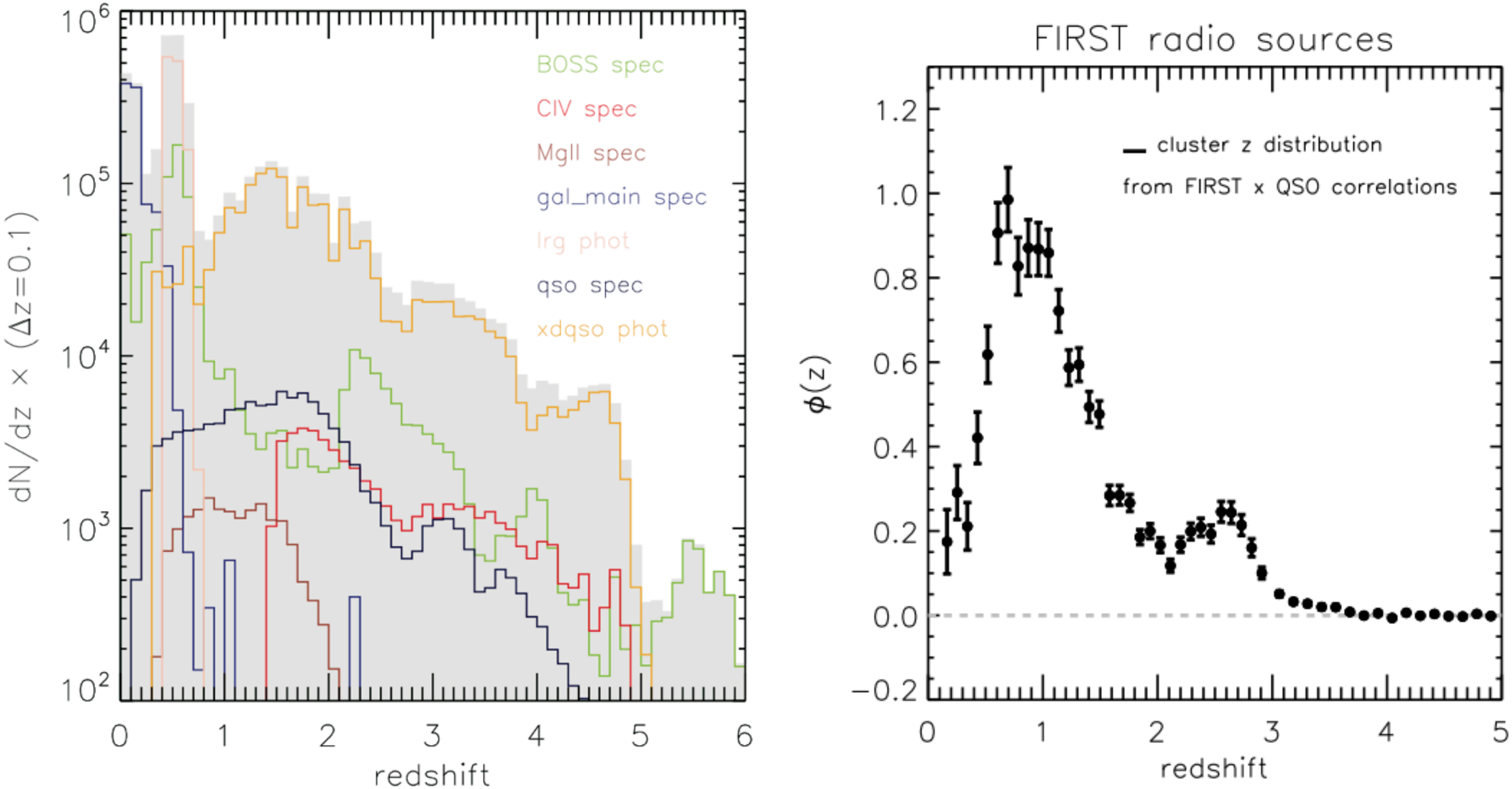
• McQuinn & White (2013): Application of optimal estimators to cross-correlation analysis



• Makes maximum use of information on linear scales, avoids integral constraint error

• Obtain errors 2-10x smaller than Newman 2008 / Matthews & Newman 2010

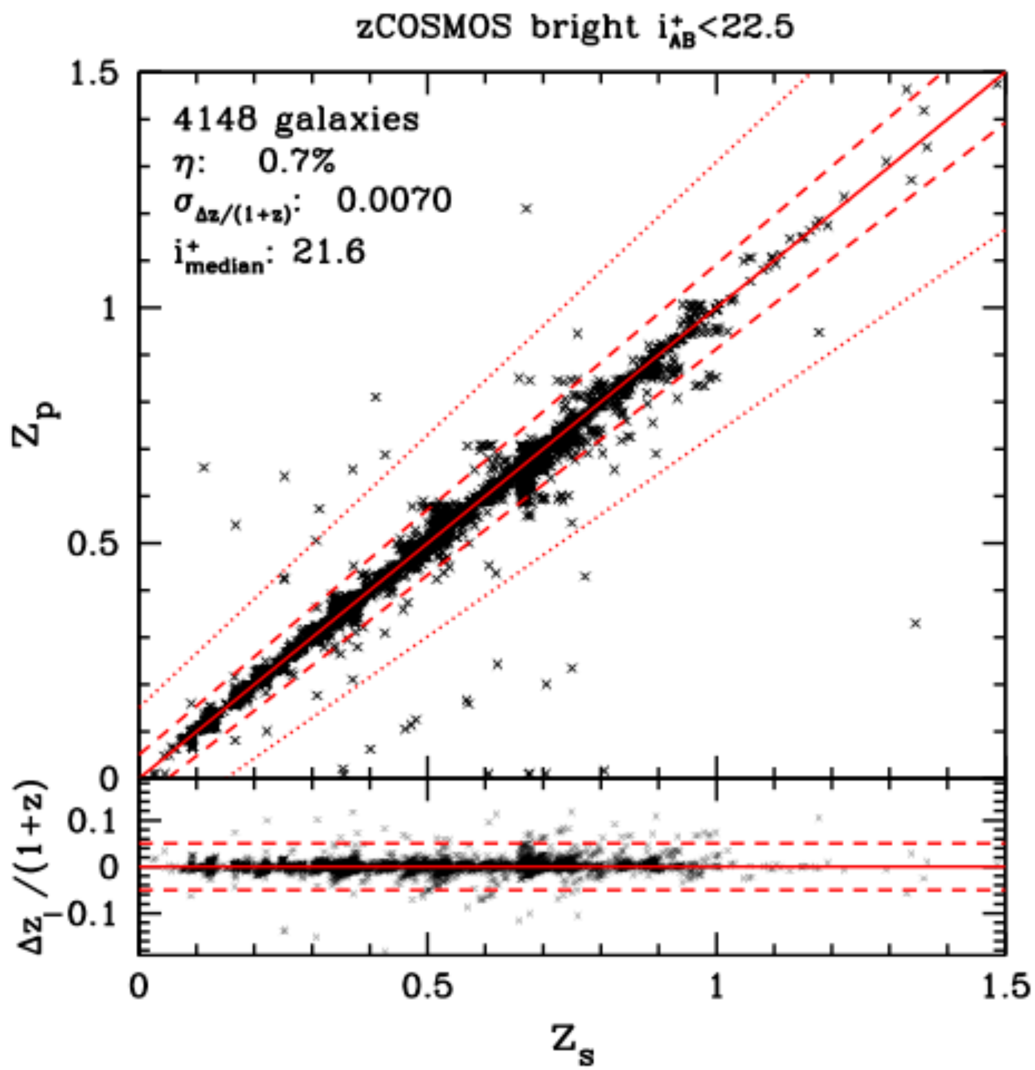
QSO samples are very useful at z>1: eBOSS and DESI will provide many



Menard et al. 2013

Current state of the art photo-z's can achieve $\sigma_z < 0.01(1+z)$ for bright objects in best case

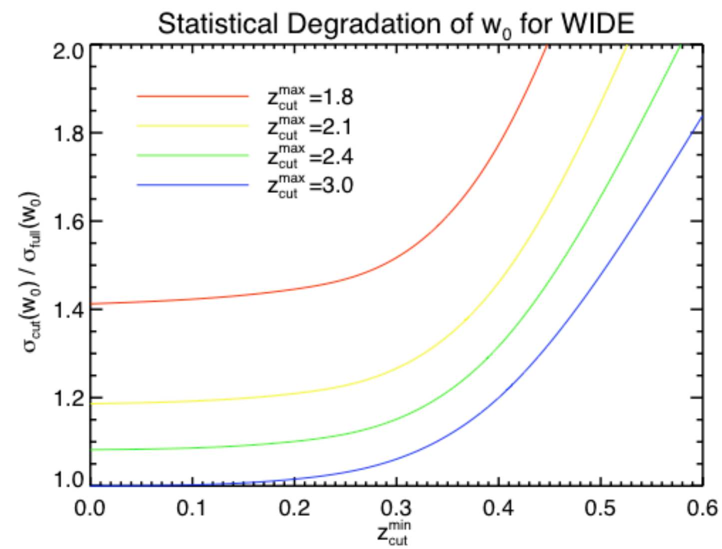
- Deep imaging in 30 bands; ≈ very low-resolution spectrum
- Predicted errors become much worse, 0.04-0.14, past z~1.25 (degenerate redshift solutions when 4000Å break passes to infrared)
- MKIDS Giga-z performance would be ~comparable if works as planned

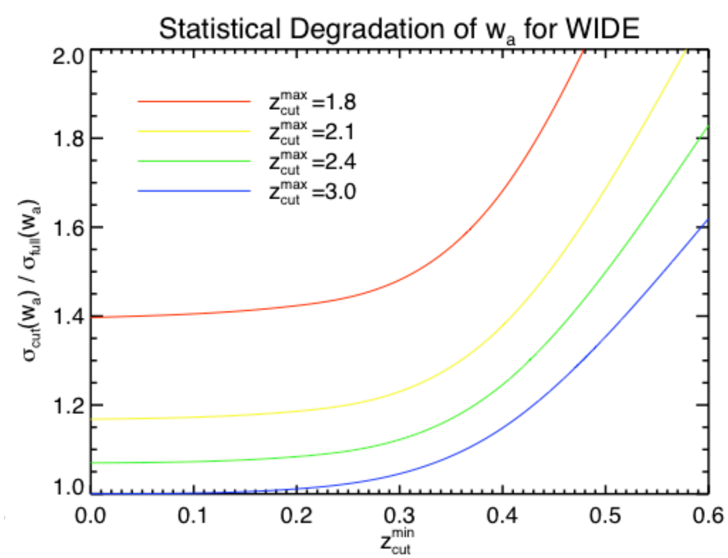


Ilbert et al. 2009

Effect of rejecting objects with particularly low or particularly high photo-z's

- **Can mitigate catastrophic outlier impact by throwing out objects with photo-z's in problematic ranges**
- **Plots at right: weak lensing error degradation (vs. random errors only) as change minimum redshift (x axis) and maximum redshift (different-colored curves)**





Hearin et al. 2010

What qualities do we desire in our training sets?

- **Sensitive spectroscopy of faint objects (to *i*=25.3)**

  - Need a combination of large aperture and long exposure times; >20 Keck-nights (=4 GMT-nights) equivalent per target, minimum

- **High multiplexing**

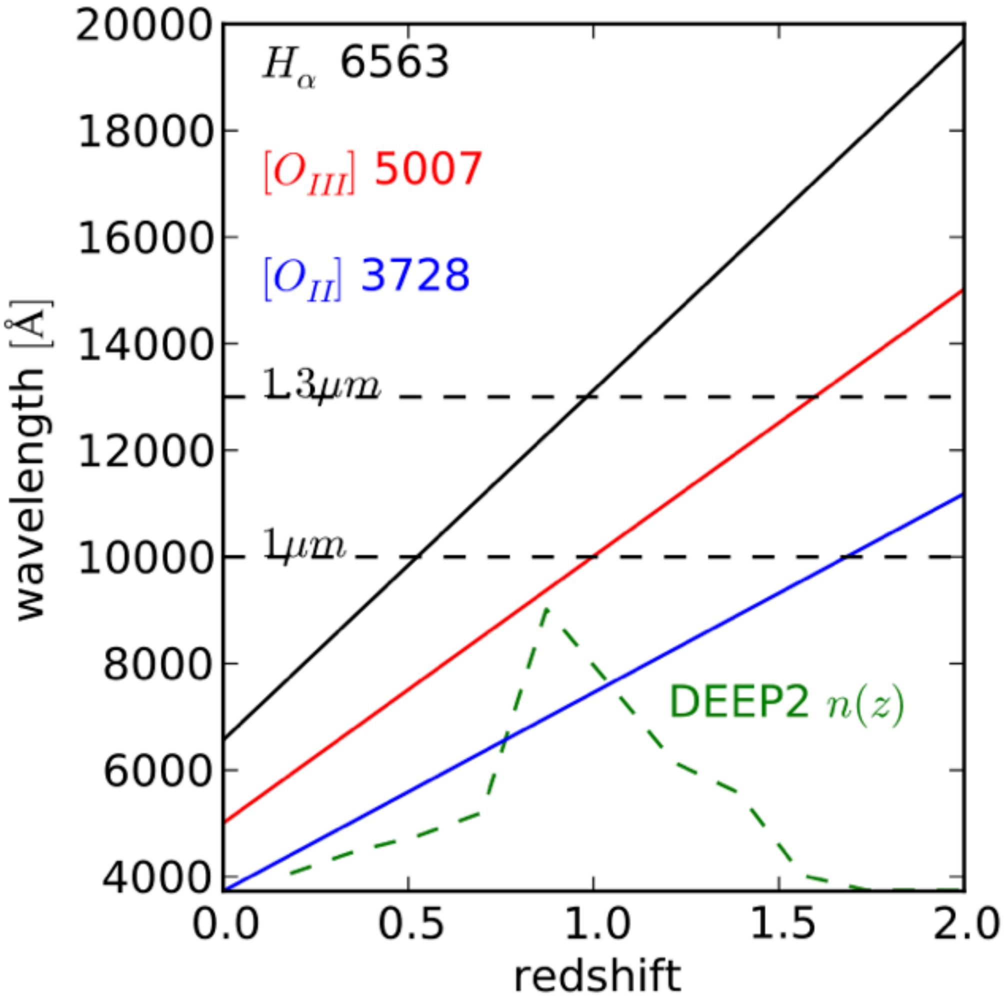
  - Obtaining large numbers of spectra is infeasible without it

What qualities do we desire in our training sets?

- **Coverage of full ground-based window**

  - Ideally, from below 4000 Å to ~1.5µm

  - Require multiple features for secure redshift

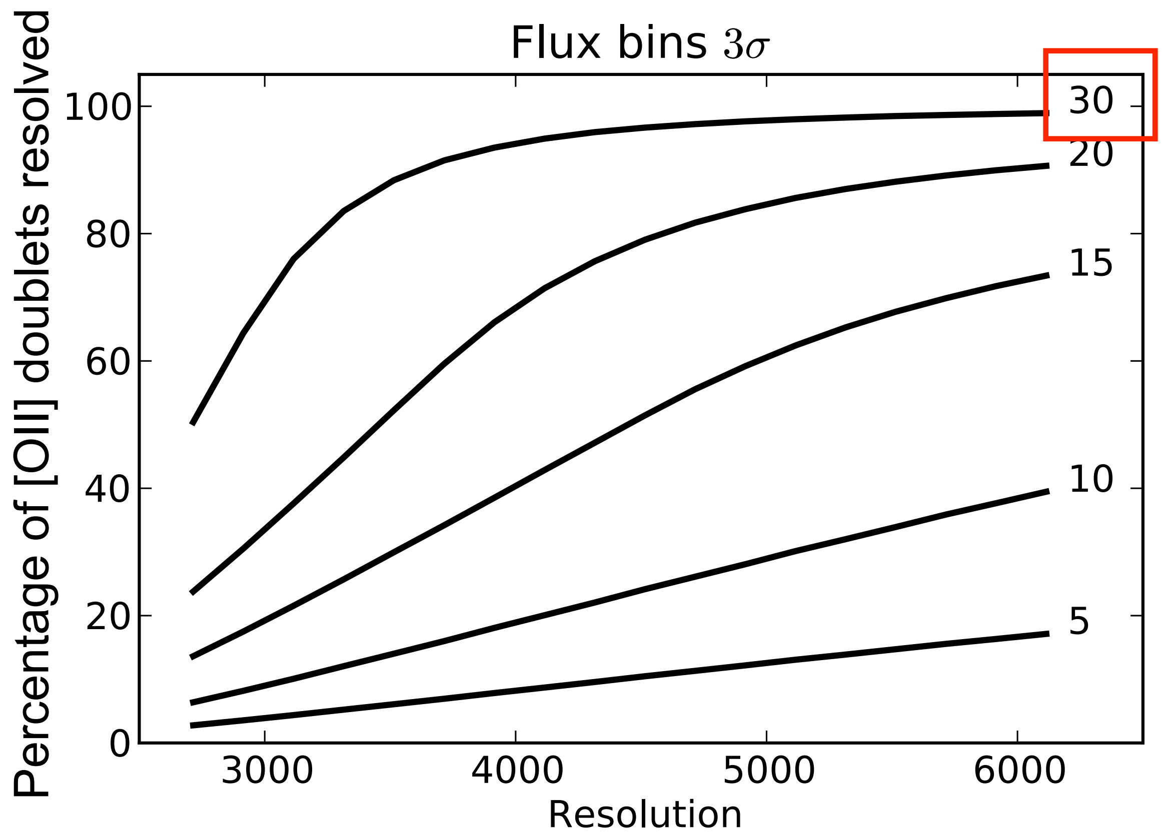


Comparat et al. 2013, submitted

What qualities do we desire in our training sets?

- **Significant resolution (R>~4000) at red end**

  - Allows redshifts from [OII] 3727 Å doublet alone, key at *z*>1



Comparat et al. 2013, submitted

What qualities do we desire in our training sets?

- **Field diameters > ~20 arcmin**

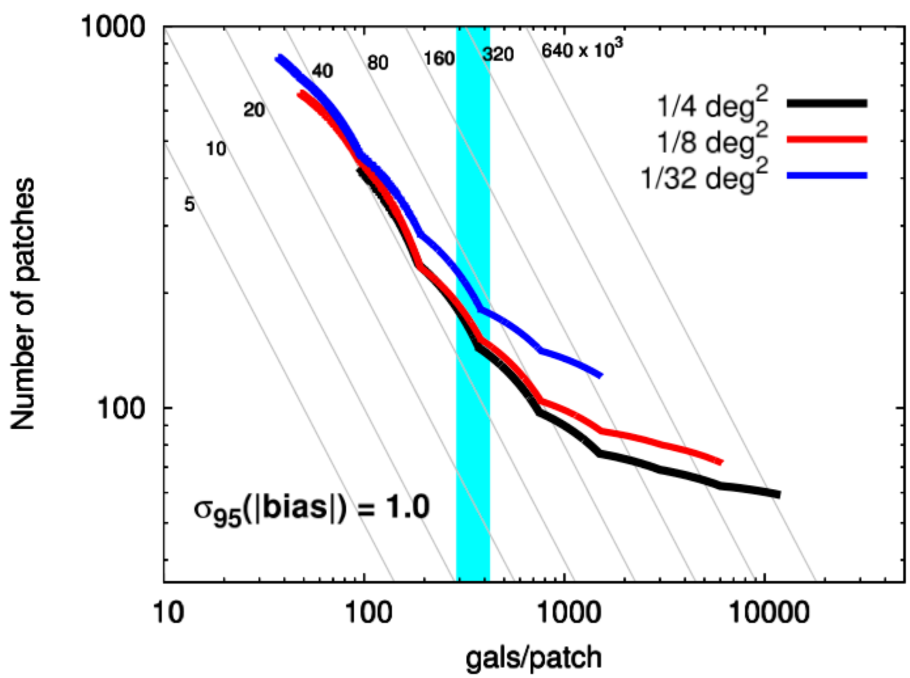
  - Need to span several correlation lengths for accurate clustering measurements (key for galaxy evolution science and cross-correlation techniques)

  - $r_0 \sim 5\ h^{-1}$ Mpc comoving corresponds to ~7.5 arcmin at z=1, 13 arcmin at z=0.5

- **Many fields**

  - Minimizes impact of sample/ cosmic variance.

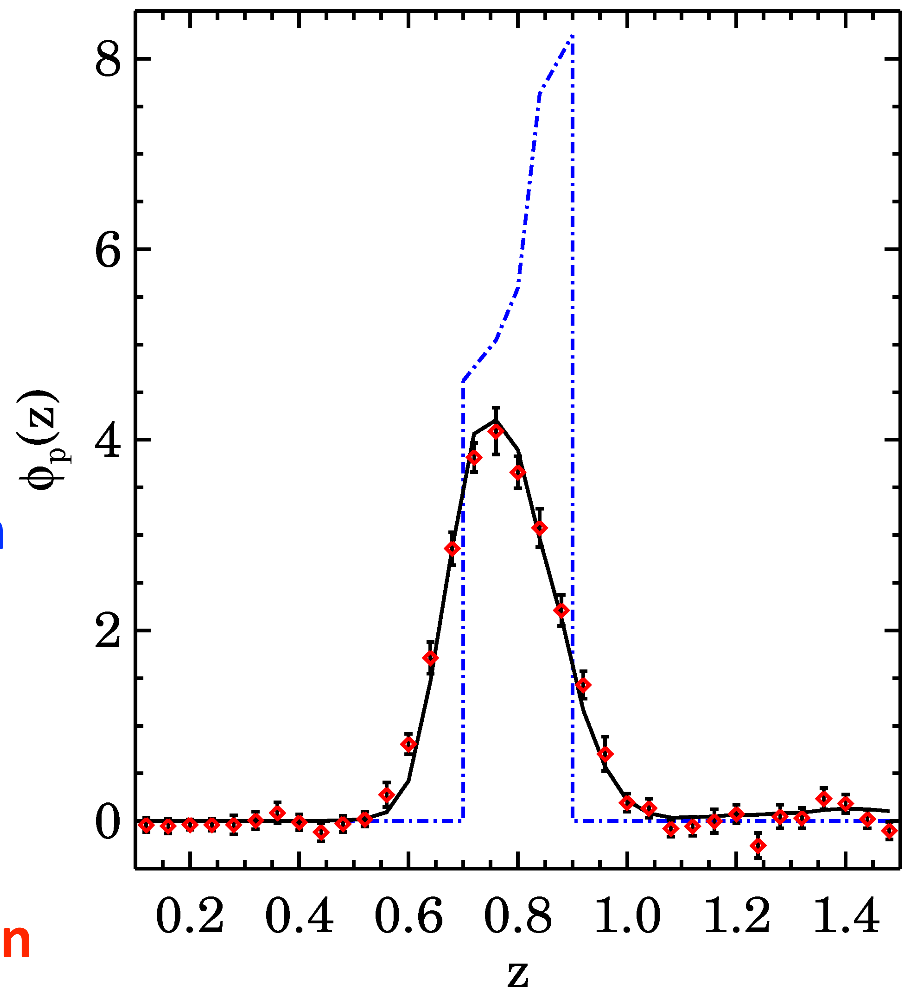
    - e.g., Cunha et al. (2012) estimate that 40-150 ~0.1 deg^2 fields are needed for DES for sample variance not to impact errors (unless we get clever)



Cunha et al. 2012

Higher-resolution information can be obtained by cross-correlating with spectroscopic samples

- Key advantage: spectroscopic sample can be systematically incomplete and include only bright galaxies
- See: Newman 2008, Ho et al. 2008, Matthews & Newman 2010, 2011



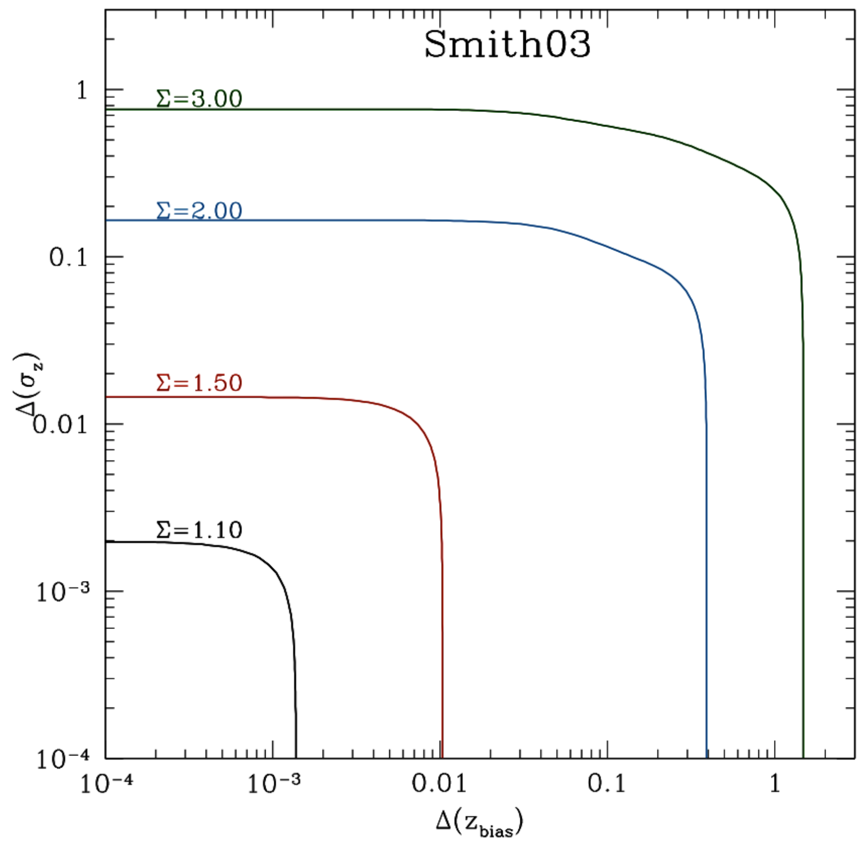
Blue: z_{phot} distribtion of objects with $0.7 < z_{phot} < 0.9$

Black: True z distribution of sample, spanning 24 widely-separated fields

Red: Cross-correlation reconstruction with only a R<24, 4 deg^2 survey

DE systematic errors from uncertainty in photo-z calibration

- Estimates based on Gaussian error models: photo-z bias, $\delta_z = \langle z_p - z_s \rangle$, and uncertainty in scatter, $\sigma(\sigma_z) = \sigma(\mathrm{RMS}(z_p - z_s))$, must be below ~0.003 - 0.01 for photo-z systematics to be subdominant in lensing/BAO (looser requirements come from better P(k) predictions)
- More realistic: need to consider catastrophic, non-Gaussian outliers. Can't be eliminated (e.g. HST shows 2% of faint DEEP2 objects are blends)

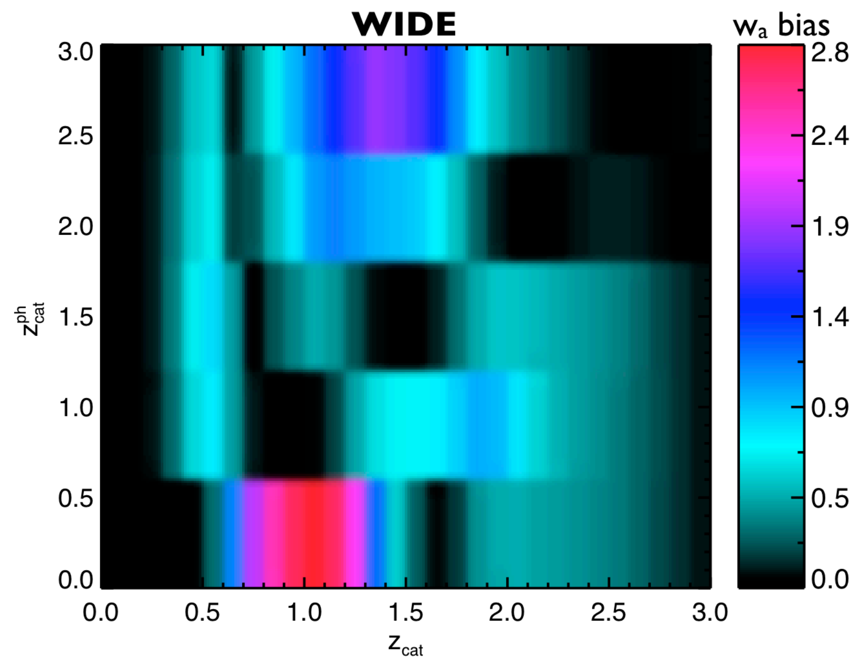


Hearin et al. 2010

- If drop all galaxies with z<0.3 or z>2.1, random lensing errors only 20% worse , but systematics much less (Hearin et al. 2010)

Systematic errors from photo-z catastrophic outliers

- More realistically: need to consider catastrophic, non-Gaussian outliers
- Can't be eliminated entirely:
  - ~2% of DEEP2 targets were actually galaxies at different *z* blurred together from ground
  - Can be difficult to distinguish one spectral break from another: degeneracies
- Some sorts of catastrophic errors worse than others
- If drop all galaxies with z<0.3 or z>2.1, lensing errors only 20% worse (Hearin et al. 2010)



Hearin et al. 2010